DOCUMENT RESUME

ED 043 657 TM 000 066

AUTHOR Roberts, Elsa
TITLE An Evaluation of Standardized Tests as Tools for the Measurement of Language Development.
PUB DATE May 70
NOTE 17p.

EDRS PRICE EDRS Price MF-$0.25 HC-$0.95
DESCRIPTORS Kindergarten Children, *Language Development, Language Programs, *Language Tests, Linguistic Patterns, *Measurement Instruments, Measurement Techniques, Nonstandard Dialects, Preschool Children, Program Effectiveness, Reading Comprehension, *Speech Skills, *Standardized Tests, Syntax, Vocabulary Development, Vocabulary Skills, Word Recognition
IDENTIFIERS Illinois Test of Psycholinguistic Abilities, ITPA, Metropolitan Readiness Tests, MRT, Peabody Picture Vocabulary Test, PPVT, Wechsler Preschool and Primary Intelligence Scale, WPPSI

ABSTRACT

Four tests--PPVT, ITPA, MRT, WPPSI--commonly used to measure language development in young children are evaluated by four criteria: (1) what development aspects do they claim to tap; (2) what do they actually tap; (3) what linguistic knowledge is presupposed; (4) what special problems face a non-standard English speaker. These tests are considered inappropriate because they fail to control question structure, to consider structures and operations the children may not have acquired, to account for dialectical differences, and to test adequately specific aspects of language acquisition. They do, however, measure the assimilation of a particular set of semantic associations and cultural values, and of a particular verbal style. It is suggested that linguistic factors be considered in all tests for young children. More research is necessary on the types of structures and operations acquired by age five and on the nature of cross-dialectal comprehension. Until the results of such research are available scores on standardized tests must be used and interpreted very carefully. (PR)

ED043657

An Evaluation of Standardized Tests as Tools for the Measurement of Language Development

Elsa Roberts

Language Research Foundation
and
Northwestern University

May, 1970

1.0 Three basic types of standardized tests have been used for evaluating language development in pre-school and kindergarten age children: intelligence tests (e.g. Stanford-Binet, WPPSI); tests designed to measure particular aspects of language abilities (e.g. Peabody Picture Vocabulary Test [PPVT], Illinois Test of Psycholinguistic Abilities [ITPA]); and readiness tests (e.g. California Readiness Tests, Metropolitan Readiness Tests). Most of these tests were designed to be used by teachers to predict the school performance potential of students, to evaluate progress or to diagnose learning difficulties; only the ITPA was designed to measure language development per se, although the others include 'language' subtests. Such standardized tests have recently been used to measure the success of pre-school language intervention programs (Cicirelli, 1969) and it is their use for this purpose which is evaluated here. In this analysis, I will show that the component aspects of language development are not isolated or controlled in the standardized tests. I will focus on four of the most commonly used tests which are representative of the major test types--PPVT, ITPA, (both language tests but different in form and content), WPPSI, and Metropolitan Readiness tests--and will ask four main questions about these tests: 1) What aspects of language development do these tests claim to tap? 2) What aspects of language do they actually tap? 3) What kind of linguistic knowledge do they presuppose? 4) What special problems does a speaker of a non-standard dialect of English face in taking these tests? This discussion may give insights into the causes of differential test performance by children of different ages and different linguistic backgrounds.

2.0 Summary of Test Contents

2.1 WPPSI: This test consists of eleven subtests: six Verbal and five Performance subtests. Only the Verbal subtests will be examined here. The WPPSI test was designed to be given on an individualized basis, with one teacher administering it orally to one child of pre-primary or early primary age. The following is a breakdown of the contents of the various subtests:

i) Information--this subtest consists of a series of content questions, e.g., "Tell me your last name", "What is the color of rubies?" There are no yes-no questions. Specific information is demanded and the responses are to be spontaneous--children are not provided with choices.

ii) Vocabulary--this subtest consists of open-ended questions, e.g., "What does X mean?"

TM 000 066

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

iii) Arithmetic--this subtest consists of story problems with visual aids, e.g., "Which is the biggest pen?", "Which two bowls have the same number of cherries?"

iv) Similarities--this subtest consists of sentences with blanks to be filled in by the child, e.g., "You ride in a train and you also ride in a ____?", "Milk and water are both good to _____?"

v) Comprehension--this subtest consists of conversational questions of various forms which the child answers to show his 'comprehension,' e.g., "Why do you need to wash your face and hands?", "Why should you go to the toilet before going to bed?"

vi) Sentence--this subtest consists of 10 sentences which the child is to repeat verbatim after the tester. It does not enter into score tabulation.

2.2 IPTA: This test contains 12 Subtests designed to test psycholinguistic processes as they are characterized by the test writers; it was originally designed as a diagnostic tool for use with abnormal children.

The processes which are tested are: Encoding, Decoding, and Association. The Channels which are isolated are: Auditory, Visual, Motor and Vocal.

i) Auditory Reception--in this subtest children are presented with yes-no questions and responses need not be verbal, e.g., "Do boys play?", "Do chairs play?", "Do chairs eat?"

ii) Visual Reception--in this subtest, a stimulus picture is shown and the child chooses one of four other pictures which is like it.

iii) Auditory-Vocal Association--in this subtest, the child is presented with verbal analogies of increasing difficulty: one well-formed sentence is followed by a sentence with a blank, e.g., "I cut with a saw; I pound with a ____?"

iv) Visual-Motor Association--the child is shown a picture and asked to point to another which goes with it, e.g., "If this goes with this, then what goes with this?"

v) Verbal Expression--the child is shown four familiar objects and told to talk about them, e.g., "Tell me all you know about this." (This is a red book.)

vi) Manual Expression--here the child is shown a picture and asked to demonstrate how the object in the picture is used, e.g., "Show me what we do with a hammer."

vii) Grammatic Closure--the child must provide the correct standard English form in a sentence where something has been omitted, e.g., "Here is a woman; here are two ____."

viii) Auditory Closure--in this subtest a record is played in which sounds are missing and the child is asked what is being said. e.g., ele/ ant.

ix) Sound Blending--words are spoken with internal breaks between sounds and the child is asked to say the word. e.g., s-a-d.

x) Visual Closure--in this subtest the child is shown a picture with partially hidden objects and is asked to find as many of a given object as he can within a limited time.

xi) Auditory-Sequential Memory--the child is asked to repeat a series of numbers after the tester has given them.

xii) Visual-Sequential Memory--child is exposed to a set of geometric figures in a particular order and asked to rearrange them in that order after they have been scrambled.

2.3 PPVT: This is an orally administered test which has been used as a diagnostic test or an intelligence test. It was designed to measure word comprehension only. The child is shown four pictures while the tester says a word. The child then chooses the picture that corresponds to the word. The score sheet indicates that response style as well as vocabulary range is considered in diagnosis (although not in the actual scoring). The child is rated for: rapport, guessing, speed of response, verbalization, attention span, perseverance, attentiveness, and need for praise.

2.4 Metropolitan Readiness: This test contains 6 subtests designed to measure readiness for school work, level of achievement, discrimination, and coordination. Only verbal tests are considered here.

i) Word Meaning--this is a test of comprehension rather than usage; it is like the PPVT in form.

ii) Sentence--this test is similar to the Word Meaning test except that the child must pick a picture which corresponds to a whole sentence or several sentences, e.g., "You would put a letter in this and mail it."

iii) Numbers--in this test the child is required to pick a picture which corresponds to the test question. Relational notions, number recognition, addition and subtraction are tested via story-problem questions, e.g., "Mark an X on the biggest apple.", "On the box where the ducks are, put a mark on 56.", "Suppose I had 3 buttons and somebody gave me 2 more; put a mark on as many buttons as I would have then."

3.0 Discussion of Test Form: In any test a child must do two things: he must comprehend and he must produce. Comprehension involves the literal comprehension of the test question; in addition it involves comprehension of the task which is demanded. Literal comprehension involves comprehension of phonological sequences, syntactic structures, lexical items, and sentence meanings (which include the comprehension of the presuppositions and implications of the question). The child's interpretation of the test question must match exactly the reading which the test writer assigned to the question.

3.1.0 The input to the child may be verbal or non-verbal. If it is verbal, it may demand a paradigmatic response or a syntagmatic response. The former type is either a word or a sentence for which the child is expected to provide some kind of equivalent substitute form (e.g., a synonym or a paraphrase). The questions which demand a syntagmatic response are either questions which are left incomplete, where the child is expected to fill in the omitted word or words, or they are complete sentences for which the child is expected to produce a sentence which follows it logically. The task required by 'fill in the blank' questions is a special one which rarely occurs in natural oral language. In these sentences the children must assign a structural description on the basis of the elements of the sentence which are not omitted; they must take semantic and syntactic cues from the sentence, extract the redundancies, and on the basis of this analysis decide which category or categories are missing from the sentence and what specific lexical item or items are semantically possible in the sentence. Non-verbal input is usually in the form of pictures or objects which the child is expected to define, label, or discuss. In this case, the child is required to switch from a visual to a verbal mode.

3.1.1 In addition to comprehension of the questions themselves, the child must understand the specific directions which accompany each subtest; further, he must understand the kind of response which is expected. For this kind of comprehension the child must be familiar with the socio-linguistic norms of the tester. It has been claimed (Baratz, personal communication) that differences in task comprehension account for the major differences in performance between groups of children from different cultural backgrounds.

3.1.2 Once the child understands the question and the task, he must then produce the desired response. There are two variable aspects of response There are two variable aspects of response type in these tests: 'verbal-ness' and 'open-ness'.

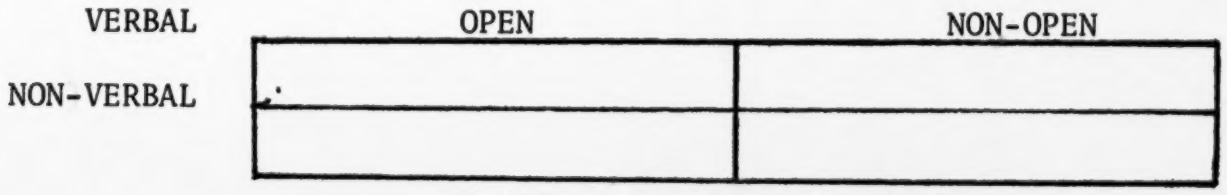

Chart I: Response Types

Any test response may be classified according to these two dimensions. Open questions are those where the child produces his own response using his particular linguistic knowledge; non-open questions are those where the child is provided with a choice by the tester.

Verbal responses require selection of a word or sentence which fits the question, or they require production of an utterance. The utterance which the child produces must have the expected informational content, sociolinguistic characteristics, and linguistic form. The child must produce a form which is grammatical, meaningful, and appropriate according to the standards of the testers. Non-verbal responses require pointing, nodding, or in some cases, gesturing and acting out (ITPA: Manual Expression).

The ITPA and WPPSI make use of the open verbal responses more than the other tests; both of these tests are administered on a one to one student to teacher basis and consequently require more active individual verbal response from the child. Responses are rated according to general norms outlined in the handbooks for teachers. In the Verbal Expression subtest of the ITPA, children are asked to talk spontaneously about simple objects presented to them. They are told to say all they can about each object. The responses are rated according to the amount and nature of verbal output. The child is expected in this test to see the object as a type of object rather than a specific object (e.g., if the block has a scratch on it, the child is denied credit for pointing this out). One final test form is used in the WPPSI which is unlike those mentioned above. This is a sentence imitation test where the child is asked to repeat verbatim sentences which are presented to him.

3.2 All of the above techniques are used to some degree in the study of language acquisition. Acquisitionists make use of verbal and non-verbal input to the child; they study response-types which vary in verbal-ness and open-ness. They employ imitation tests to measure linguistic competence of children. There are, however, three important ways in which their methods of assessing language development differ from those of the standardized testers.

3.2.1 Acquisitionists use the tests to learn about the language of children rather than to fit children into predetermined categories. They design tests for the purpose of gaining insights into the developing linguistic system, rather than for the purpose of ranking children according to prescriptive norms. Acquisitionists are interested in children's mistakes insofar as these mistakes give insights into the mental processes of the children; thus, error analysis is an important tool of acquisitionists, while it plays little role in standardized testing.

3.2.2 Acquisitionists control the linguistic content of the tests very carefully: a) They test specific hypotheses about particular structures of operations rather than general undefined notions of 'vocabulary', 'comprehension' and 'meaning'. b) They only use structures which are known to be within the competence of the tested children unless the structures are the target of the testing. c) They are careful to eliminate semantic cues which might provide the child with redundant information that helps him to respond correctly without actually understanding the tested structure.

3.2.3 Acquisitionists do not rely solely on test situations to assess language development. There is a strong tradition of observational study of children using language in natural conversation settings. Tests are used only to assess very specific aspects of language acquisition. In addition, all available evidence indicates that language used in test situations is qualitatively different from spontaneous language used in natural settings.

4.0 Discussion of Test Contents

Two substantive areas of language acquisition are explicitly tested in these four standardized tests: vocabulary and syntax. We will discuss the adequacy of these tests as measures of the development of vocabulary and syntax. In addition, we will discuss the kinds of linguistic knowledge which are presupposed in the verbal subtests of all the tests. All the tests include vocabulary subtests. Only the ITPA and WPPSI have subtests which might be considered tests of syntax. In addition the ITPA has two phonological tests which I will not discuss here.

4.1 Vocabulary Tests

There are three main ways in which the vocabulary tests are inadequate: i) They tap only semantic information without measuring the child's knowledge of syntactic information associated with test items; ii) the syntactic knowledge which is demanded, while it is not related directly to the test item, is complex and not controlled; iii) only one grammatical category is tested; however, the presentation of this category is often ambigous and therefore potentially confusing. These points are elaborated below.

In all of the vocabulary subtests, 'knowing' a word is equated with having a particular semantic association with the word, in the form of a pictorial image or another word. Some semantic property of the tested word must be related to one property of the answer. Thus, knowing that a goose is a bird is sufficient; it is not also necessary to demonstrate knowledge of the fact that goose is an animate count noun which has a suppleted plural form rather than a regular plural. No attempt is made to find out if the child understands the usage of items in sentences. For example, tell and promise differ not only semantically but syntactically as well; in the

following sentences the one who leaves is in one case Bill and in the other Henry:

a. Bill promised Henry to leave.
b. Bill told Henry to leave.

The difference in syntactic properties of these verbs must be part of the mature speaker's knowledge of these items. This kind of knowledge is not tapped.

In one test the following sort of questions are found:

To sparkle means to ____. (attempt, command, shine)
We go to school to _____. (learn, sing, travel)

In order to choose correctly the child must understand the structures of the question; he must be able to extract semantic cues from the sentence. However, in no case is he required to know the syntactic properties of the verbs which go in the blanks. He does not need to know that sparkle cannot occur with an object, that attempt occurs with a sentence complement, that command can occur with or without an object and that shine (unlike sparkle) can also occur with or without an object. To answer correctly the child needs to have the syntactic structures of the questions in his competence but he is not required to show knowledge of the snytactic properties of the vocabulary items which fit into the blanks.

To take one further example, knowledge of the features of nouns is not tested; in the PPVT, the word cash, which is a mass noun, is followed by the count noun, whale; the child is given no opportunity to demonstrate that he does or does not have the count/mass distinction in his competence. He is not required to show that he knows mass nouns can be preceded by some but singular count nouns cannot.

The above examples illustrating syntactic properties of the verbs promise and tell, and illustrating the count/mass distinction are particulary important because recent studies have shown that these aspects of English are not acquired until quite late. (Chomsky, 1968; Hatch, 1969). Thus, real developmental differences are signalled by differential knowledge of these structures.

The Auditory Reception subtest of the ITPA does tap a specific kind of lexical information in a controlled way although the test-makers seem to be unaware that they are doing so and in fact, claim to be testing a much more general aspect of language: "the ability of the child to derive meaning from verbally presented material." (ITPA Manual, p. 11-12). In this test, the child is presented with a yes-no question of the following sort: "Do chairs eat?", "Do chairs play?", "Do boys play?". In order to give the correct answer to the above questions it is necessary to know:

i) that chairs are inanimate
ii) that boys are animate
iii) that verbs eat and play require animate subjects.

The specific properties of the nouns and verbs, and their co-occurrence relations are tested here. Mixed in with these questions are others in which a different kind of knowledge is tested: e.g., "Do dogs fly?" In the "dog" sentences, it is knowledge of the world which is being tapped. To answer, the child must know that dogs do not have wings and special apparatus (e.g. wings) is necessary for flying.

In all the vocabulary tests, a high proportion of the tested items are nouns. This is a natural outcome of the method of testing; word associations or matchings of words and pictures naturally are drawn from items with physical

characteristics or with one-word synonyms.

The categories which are tested in each vocabulary subtest are given in Chart II. The class noun refers only to those items which are completely unambiguous as to classification (e.g., cat, nuisance); the noun/verb class refers to words which are either nouns or verbs, depending on the context: nominalizations are those verbs which occur with an ing ending (e.g., knitting, skiing); undeclined noun/verbs are those which occur without an ending (e.g., test, gamble, nail). Some items occur out of context in the test. Thus, the child is required to determine the part of speech in addition to the semantic content of the word; this task is complicated in the cases where the item is categorically ambiguous (e.g., knitting, nail). It is interesting to note that in some cases the examiner may unknowingly aid the student in ascertaining the categorical features of the item by the form of the question he poses. Thus, in the WPPSI subtest the examiner is instructed to say either, "What does X mean?: or "What is a X?" In case the examiner chooses the second question form, he may give the student information about the count/mass distinction; the determiner a can be used with singular count nouns only. Without this determiner, the number marking of the verb tells the student if the noun is a plural count noun or a mass noun. In like manner, this question form eliminates the ambiguity of noun/verb types. Thus some students may be presented with fewer choices about categorization because of the wording of the question.

	WPPSI (Vocabulary) 22 items	PPVT (first 50 items)	Metro (Word Meaning) 16 items	ITPA (Manual Expression) 15 items
Noun				
count	54.5%	64.0%	81.25%	100%
mass	9.0%	4.0%	____	___
total	63.5%	68.0%	81.25%	100%
Noun/Verb				
nominalization		18.0%	6.25%	___
unde-clined	27.5%	14.0%	____	___
Verb	4.5%	____	____	___
Adjective	4.5%	____	____	___
	100%	100%	100%	100%

Chart II: Vocabulary Items

4.2.0 Syntax Tests

There are three main criticisms of the sentence tests: i) the specific linguistic tasks which they test are trivial and systematically biased against speakers of non-standard dialects of English; ii) the tests do not take factors of processing complexity into account; iii) the tests do not take developmental factors into account.

4.2.1 The ITPA grammatic closure test is specifically designed to measure particular syntactic structures. There are 33 items on this test; of these, 24 items may have different forms in some dialects of English. In order to be correct, the answers must be in standard English. Adequate performance on this test requires nothing less than ability to produce SE plural, possessive reflexive and negative constructions. For example, there is a plural question where the child must supply the form children; if he supplies chilluns instead, his answer is to be marked incorrect. In another question, the child is to change a sentence with some to its negative counterpart. A sentence like, "I have some eggs" is to be changed to, "I don't have any (eggs)"; if the child says, "I don't have none" or "I don't have no eggs", which are the grammatical counterparts in Black English, his answer is incorrect again. The testers, however, claim to be measuring "ability to make use of the redundancies of oral language...in acquiring automatic habits for handling syntax and grammatical inflections..." (ITPA, p. 13) The aspects of syntax which are tested in this subtest are limited to superficial morphological structures (plural formation, possessive endings, comparative endings). There are a few sentences where the child is asked to produce a sentence with a structure different from the structure of the cue sentence:

1. The boy is writing a letter; the letter has been _____
2. The boy likes to play; the boy is _____
3. The boy has some food; the boy doesn't have ____

Here again, the testers do not isolate the structures they are trying to test. The two sentences of (1) differ in tense and in voice (active vs. passive). One of the sentences of (2) contains a complement sentence while the other does not; the derivational history of these two sentences is quite different. Only the sentences of (3) are closely related derivationally. We do not wish to make any psycholinguistic claims about the reality of linguistic derivations for sentence comprehension or language acquisition, but do wish to point out that the sentences used in this subtest of the ITPA are limited in number and type of syntactic structure. Moreover, there is no apparent linguistic motivation for the selection of these particular syntactic structures other than to test competence in standard English.

4.2.2 The Sentence subtest of the WPPSI might also be construed as a test of knowledge of syntactic structures. In this test children are asked to repeat sentences verbatim. Psycholinguists agree that the imitatability of sentences depends to some degree on the subject's ability to process the input sentence; and this processing ability is related to the level of development of the subject's linguistic competence. The ease of processing may depend on any number of factors: however, there is no consensus on the exact nature of the processing task. It has been suggested that the number of contentives (semantically loaded words) per noun phrase and their structural relations to each other are crucial factors in sentence imitation tasks. (Smith, 1970) According to this analysis, the noun phrases of the last sentence

are less difficult to process than those of the first four sentences. The first two sentences each have one noun/phrase with two contentives; the third has one with two contentives and one with three contentives; the sixth sentence has two noun phrases with two contentives each; the seventh has a noun phrase with five contentives, and so on. There is no correlation between the number of contentives, per noun phrase, the number of complex noun phrases per sentence and the order of the sentences in the list (which presumably goes from easy to difficult).

An alternate explanation of J.J. Fodor and M. Garrett (reviewed in Smith, 1970) attributes complexity in noun phrase processing to the number of underlying sentences. There is consensus among psycholinguists, that ease of repetition depends on more than sentence length; it is somehow related to the structural characteristics of the sentence and the level of linguistic competence of the child. In the WPPSI sentence subtest, sentence length is the only factor which is systematically varied. Each sentence is longer than the previous one by one or two word units. Differences in syntactic structures are not taken into consideration. In fact, so little attention is paid to syntax and grammar that the first sentence of the test is "My house" which hardly qualifies for sentencehood by anyone's criteria. Two other adjacent sentences further down in the test are:

1. It is very nice to go to camp in the summertime.
2. Peter would like to have new boots and a cowboy suit.

These sentences have very different structures in terms of numbers of embeddings, complements, and kinds of syntactic transformations applied. The following sentences (which do not occur in the WPPSI) are different in length but similar in structure.

1. The cat likes fish, liver, horsemeat, pork, and chicken.
2. The cat likes liver, shrimp and chicken.

In these sentences, the memory factor is the major variable. The above examples serve to illustrate that the imitation test could have been used to measure either linguistic competence or memory; however, as it is presently set up the test does not isolate either kind of variable.

4.2.3 The sentence tests of all the standardized tests fail to take into consideration the level of language acquisition of the test-taker; these tests presuppose virtually full adult competence. Recent studies have shown that children have not acquired some adult structure-types until age ten. (C. Chomsky, 1968; Hatch, 1969). The kinds of structures which a child must have in his competence to fully understand the test questions are examined below. We do not wish to claim that successful test performance is totally determined by knowledge of these structures; the relative importance for comprehension of knowledge of syntactic structures and ability to extract semantic cues is not known. It may well be that children depend critically on semantic information within the test question, particularly in closed questions where they are required to choose among four pre-determined answers. However, all studies of language acquisition to date indicate that the ability to comprehend sentences is determined to a considerable degree by the ability to comprehend their structural characteristics.

Certain structures which occur in the tests are known to be beyond the comprehension level of most five year old speakers of standard English. In addition to these, structures which have not been investigated by acquisitionists but which are felt to be potential sources of difficulty will be examined

here. The following have been shown to be structures beyond the competence of most kindergarten and pre-kindergarten children (E. Hatch, 1969):

i) Be-passives: children understand (and produce) sentences using the got passive before they understand the corresponding be passive sentences, e.g., 'the dog got hurt' is acquired before 'the dog was hurt'; the agentless forms are acquired before the forms with agent: e.g., 'the dog (got, was) hurt by the cat (Bates, 1969).

ii) Relative clauses with passives; although relative clauses are acquired by age four; sentences of the form: 'the window which (was, got) broken is over there' are probably not mastered until later since the passive is not mastered until then.

iii) Time connectives: in sentences where the surface grammatical order of conjoined sentences is different from the temporal order of the events, kindergarten children have comprehension difficulties. Difficult sentences are:

"Do X but do Y first."
"Do X after Y."
"Before X, do Y."

Sentences which present no difficulties are:

"Do X and then Y."
"Do X before Y."

iv) Conditionals: kindergarten children have difficulty with sentences where the conditional markers if/then, if not/then, unless/then, and unless/then not are present. In addition, conditionals in complex sentences with tense differences cause difficulties.

"What would you do if you fell?"
"I wish I had a book."
"What should you do when you fall."

v) Pronominal reference: in sentences with complements, children have difficulty identifying the deleted pronoun of the complement clause, especially when the verb in question is exceptional with regard to the rules of complementation (Chomsky, 1968). Thus, the following sentences are confusing to children of kindergarten age:

"John asked Bill to leave."
"John promised Bill to leave."
"John was nice to leave."
"Tell him where to go."
"Ask him where to go."

Chart III illustrates the number of sentences in the Metropolitan and the WPPSI tests which contain each of these structures; only test questions themselves were examined. No count was made of structures used in instructions (which, of course are also crucial for test comprehension, and might be separately investigated). The ITPA test questions are not included in this tabulation since all except the grammatical closure test, discussed above, had one or two simple test frame questions into which the individual test items fit and the grammatical structures were minimally varied.

	WPPSI	Metro
Be-passives	5	3
Relatives with Passives	-	2
Time connectives	1	5
Conditionals	7	4
Pronominal Reference	2	2

Chart III

(note: in Charts III and IV, occurrences of structures in test instructions are not included; figures are given in absolute numbers because it was not deemed possible or interesting to tabulate percentages of structures.)

In addition to the above structures which have been shown experimentally to present comprehension difficulties, the following structure and operation types which occur in test questions and have not been investigated might present difficulties for kindergarten children. The examples listed with each item are taken from the tests; to answer the child must point to a picture which corresponds to the question or respond spontaneously.

i) Indirect questions:
"Mark the one which tells how many balloons there are."

ii) Various types of deletion: the deleted elements in the following examples are represented in parentheses.
a) Deleted relative pronouns:
"This animal has many things (that) other animals have."
b) Verb deletion in conjoined sentences through gapping:
"Girls grow up to be women and boys (grow up) to be men."
c) Relative clause reduction:
"There they saw an organ grinder with a monkey (who was) dressed up in a little jacket and a funny cap."

iii) Purpose clauses:
"What do you need to put two pieces of wood together?"
"What do you do to make water boil?"

iv) Comparatives:
"It is <u>better</u> to build a house of brick <u>than</u> of wood."
"The price is <u>as</u> high this year <u>as</u> it was last year."

v) Quantifiers:
"Which bowls have <u>the same number</u> of cherries?"
"<u>Each</u> boy had <u>some</u> meat."
"<u>Both</u> boys have <u>some</u> meat."

vi) Uncommon structures which are used in formal writing or speaking styles or in one regional form of colloquial speech might cause difficulties because they are uncommon in the linguistic environment of the child.
a) Fronting of the preposition along with its object in questions of relative clauses:
"From what animals do we get milk?"
b) <u>Must</u>: the use of <u>must</u> instead of have to (or some other form):
"What must you do if you fall?"

c) The thing to do:
"What is the thing to do if you fall?"

vii) Lexical items which have the same phonological shape but which have different meanings and occur in different structures: e.g., make: this verb has three distinct senses which are used in adjacent sentences of one subtest:

"What must you do to make water boil?"
"How many pennies make a nickel?"
"What is bread made of?"

In the first of these sentences, 'make' can be paraphrased as 'cause'; in the second it can be paraphrased as 'constitute'; in the third, it can be paraphrased as 'goes into'. In the first sense, it can occur in both the active and passive; in the second it can occur with active only, and in the third, with passive only.

viii) Multiple embeddings: The effect on children's sentence processing of more than one embedding per sentence has not been fully investigated. It is known that children produce sentences with fewer embeddings than adults, and it might be hypothesized that multiple embeddings cause comprehension difficulties even if the embedding structures and processes themselves have been mastered. Thus in the following sentences, the number of underlying sentences and/or surface clauses alone might be an obstacle to comprehension, particularly if this test is administered orally and the child has no recourse to re-reading or going over the sentence on his own:

"Put a mark on as many socks as the three children need to keep their feet warm."
"Mark the picture of the thing which makes it possible for you both to see and hear people who are in another city far away."
"In Switzerland the cows wear bells around their necks so the boy can find them when they wander away."

Chart IV indicates the number of times each of the above operations or structures are found in the WPPSI, ITPA and Metropolitan tests. This chart indicates again that the tests make extensive use of structures which may interfere with the comprehension of five year old children and these structures are used in an uncontrolled manner so that it is impossible to ascertain exactly what it is about any given sentence which causes difficulty.

	WPPSI	Metro	ITPA
Indirect Questions	----	7	----
Relative Pronoun Deletion	----	3	----
Gapping	2	----	----
Purpose Clauses	2	1	4

con't

Comparatives	2	10	----
Quantifiers*	4	2	----
Preposition Fronting	2	----	----
Colloquialisms	5	1	----
Make	6	----	----

Chart IV

*Questions with numbers, or with how many and how much are not included in this tabulation.

What is striking about the configurations represented on Charts III and IV is that occurrences of particular structures appear in clusters. All occurrences of the indirect question and relative pronoun deletion are found in the Metropolitan test; only the WPPSI test has instances of gapping, preposition fronting, and various uses of make. There is an appreciably greater number of uses of comparatives on the Metropolitan test than on the others; colloquialisms are used more in the WPPSI test than elsewhere. Thus, it seems that the absence of a particular structure within the competence of an individual child could affect his performance on one test or subtest quite significantly. No analysis of responses which takes sentence structure types into account is available; such an analysis might give insights into the role of the stage of language acquisition in test performance, and into differential performance on various subtests, where clusters of a single structure type are found.

5.0 Discussion of Biases against Non-standard Dialect Speakers

Finally, we turn to the question of special problems which face speakers of non-standard dialects of English. There are four areas where the tests might present additional tasks to children who do not come from a background where SE is spoken:

i) The content of the test questions and expected responses
ii) The verbal style required by the test
iii) The non-linguistic factors inherent in the test situation
iv) The linguistic aspects of the test

5.1 Substantive biases in standardized tests can include culture specific vocabulary items, culture specific pictures used in vocabulary tests, culture specific information questions, and even dialect specific linguistic questions. In these cases, the "correct" answer involves knowledge of the particular alnguage or culture of the tester.

There are two ways in whcih the vocabulary tests can be biased against children of a particular subgroup: either the object which the test word itself can be different in the dialect of the subgroup (e.g., spectacles). The absolute numbers and the percentages of potentially culture-specific (to Standard English culture) items are given in Chart V.

	number/total	percentage
ITPA (Manual Expression)	3 /15	20
WPPSI (Vocabulary)	4 /22	19.8
Metropolitan (Word Meaning)	6 /16	37.5
PPVT (first 50 items)	13 /50	26

Chart V

In the same way, an information question which presupposes a particular cultural norm reflects bias. On the WPPSI comprehension subtest, for example, the question "Why do you need to wash your face and hands?" presupposes that you do need to wash your face and hands, which may not be a cultural universal. A good response is: "to get clean" or "so you won't get germs"; a less acceptable response is: "they're dirty"; an even lower-rated response is: "Mother tells you to." (This test is one of those used to measure 'intelligence.')

On the same test the question "why are criminals locked up?" is considered well-answered if the child includes the idea that locking up criminals is a deterrant, that it is for the protection of society, for punishment, revenge, rehabilitation and/or segregation. A bad response is: "they're bad, they kill people (in the present tense)...they're dangerous." It is interesting to note that a present tense answer is explicitly given the lowest rating; in some dialects of English the past tense morpheme of standard English, ed (e.g., killed) often has no phonological realization; present and past tense forms of the verb kill are pronounced alike in these dialects. High performance on this test entails nothing less than full socialization into the culture of the dominant subgroup, the culture of speakers of the dominant dialect of English, in addition to some degree of assimilation of their dialect.

Finally, a test of grammatical forms which are not in the dialect of a speaker is all but impossible for him to do well on. The ITPA grammatic closure test is thus inherently biased against speakers of BE and it is not surprising that they do less well on this test than SE speaking children. In fact, it would be rather surprising if BE speakers characteristically performed as well as SE children on this sort of test, since this would indicate that these children are successfully performing a cross-dialectal production task, in addition to the other tasks required by the test.

5.2 The verbal style required by a test can be culture specific. For example, a standard of articulate description in one culture might be specificity and brevity. A child from this culture, describing an item in the ITPA verbal expression test might say in one short sentence that the block is red and has a scratch, thus failing to meet the prescribed criteria of expressiveness set by the test designers (namely, quantity and generality of description).

In addition, the norms for verbal interaction might be different in the speech community which the child comes from. Susan Philips has shown that children from the Warm Springs Indian Reservation in Oregon are inhibited in speech situations where they are called on by the teacher to produce a response

to a 'pop' question in front of the class (Philips, 1970). These children come from a community where they are never asked to perform by request of another person; they volunteer, or speak spontaneously and they do this only when they are certain of the desired response. Thus, the particular norms of adult-child interaction in the community that the child comes from will strongly affect his performance on individualized tests such as the WPPSI and the ITPA.

The particular standards of verbal style of Standard English culture are explicitly outlined on the individualized record sheet of the PPVT. Test behavior criteria are listed as a guide in diagnosis; these criteria include: "examples needed (only 1), type of response (subject pointed), rapport (easily attained), guessing (resisted guessing), speed of response (fast), verbalization (talkative), attention span (very attentive), perseveration (none noted), need for praise (little needed)." (The highest value for each of these areas of test behavior is noted in parentheses. These are the style norms of Standard English cultures; the child's conformance to these norms affects performance and evaluation of any standardized test and, in particular, individualized tests such as the WPPSI and ITPA, where there is constant subject-tester interaction.

5.3 Situational factors also can act against speakers of non-standard dialects. The fact of being tested itself can intimidate a child so that his performance is inhibited; the mere awareness on the part of the child that he is expected to produce according to norms not indigenous to his own culture can cause him to resist the testing by refusing to participate. In addition, forced interaction with an adult who speaks another dialect, has a different color of skin, or comes from another culture can affect test performance in the same way. Finally, the child may have difficulty in producing because he does not understand the nature of the tasks presented to him. These sociolinguistic factors will not be documented here; there is ample support for these claims in recent research. (e.g., Phillips, 1966).

5.4 Finally, the tests may be biased in subtle linguistic ways. It is not known (although the Language Research Foundation is currently conducting an experiment in this area) to what degree speakers of other dialects and in particular young speakers understand SE. There may be phonological difficulties for speakers of other dialects in the oral tests (the sort of problems the Wepman test has been used to illustrate recently; (Karger, 1970)) or there may be difficulties in understanding particular syntactic structures because of dialect differences. The semantic connotations and denotations of words, the implications and presuppositions of sentences which differ from dialect to dialect may cause difficulties. All of these are open questions and their answers bear directly on the problems of testing young children of diverse backgrounds. In any case, non-standard dialect speakers must perform two tasks which speakers of SE need not perform: i) they must decode forms from another dialect and assign meanings to these forms; and ii) they must encode into a dialect which is not their own. The exact nature of cross-dialectical comprehension and production tasks is not known. It is clear, however, that these tasks are required of SE speakers.

6.0 Conclusion

The intent of this analysis has been to show that the standardized tests examined here are inappropriate measures of language development because:

(i) they fail to control the structure of the questions; (ii) they fail to take into consideration the types of structures and operations which children have not yet acquired by age five; (iii) they fail to take into account dialect differences; and (iv) they test specific aspects of language acquisition only trivially. It appears that what these tests do measure is the degree to which a child has assimilated a particular set of semantic associations, a particular verbal style, and a particular set of cultural values. They assume homogeneity of linguistic competence (except where a trivial aspect of this competence is tested and acceptable performance is equated with production of SE forms). They ignore socio linguistic factors crucial to test performance.

It may also be the case that these tests are questionable measures of intelligence or cognitive development. If linguistic factors (level of linguistic development or dialect differences) hinder comprehension or production, then a child will be unable to demonstrate knowledge of the cognitive task in question. It may well be, for example, that a child of five has acquired the notion of causality without all the accompanying linguistic forms, particularly the SE forms. If this child were asked "What makes you cry like that?" he might be unable to answer, while he could answer a paraphrase of the same question without difficulty: "Why are you crying like that?" In an information question, where knowledge of a particular fact is being tested, a child might be unable to answer if the wording is: "Tell me whether elephants have wings" whereas he could easily answer the alternative, "Do elephants have wings?"

Thus, linguistic factors must be taken into consideration in tests for young children even if these tests are not specifically designed to test language. Until there is a great deal more research on the types of structures and operations acquired by age five, and on the nature of cross-dialectal comprehension, we must be extremely careful in how we interpret the results of standardized tests and the uses to which we put them.

References

Bates, Reed., A Study in the Acquisition of Language. Ph.D. Dissertation, University of Texas, 1969.

Calfee, R.C. & Venezky, R.L., "Component Skills in Beginning Reading", in Goodman, K.S. & Fleming, J.T. (eds.) Psycholinguistics and the Teaching of Reading. Newark, Delaware: IRA, 1969.

Chomsky, Carol S., The Acquisition of Syntax in Children from 5 to 10. Cambridge, Mass., M.I.T. Press, 1969.

Cicirelli, V. et al., The Impact of Head Start: An Evaluation of the Effects of Head Start on Children's Cognitive and Affective Development. Vols. I and II. Bladensburg, Md., Westinghouse Learning Corp., 1969.

Dunn, L.M., Peabody Picture Vocabulary Test. Form B. Circle Pines, Minn., American Guidance Service, Inc., 1959.

Hatch, E. Four Experimental Studies in Syntax of Young Children., Inglewood, Calif., Southwest Regional Laboratory for Educational Research and Development, TR 1, March, 1969.

The Syntax of Four Reading Programs Compared with Language Development of Children., Inglewood, Calif., SWRL, TR 21, 22 December, 1969.

Hildreth, G.H., Griffiths, N.L., McGauvran, M.E., Metropolitan Readiness Tests., Form A., Harcourt, Brace, and World Inc. 1965.

Karger, Trudy, The Wepman Test and Dialect Differences, Harvard Graduate School of Education, Unpublished paper, 1970.

Kirk, S.A., McCarthy, J.J., and Kirk, W.D., Illinois Test of Psycholinguistic Abilities. Revised Edition, Urbana Illinois, University of Illinois, 1968.

Philips, Susan U. "Acquisition of Rules for Appropriate Speech Usage." Paper read at 21st Annual Round Table on Languages and Linguistics, Georgetown University, March 13, 1970.

Phillips, Judith. "The Effects of the Examiner and the Testing Situation upon the Performance of Culturally Deprived Children," Phase I - Intelligence and Language Ability Scores as a Function of the Race of the Examiner, E.R.I.C., 1966.

Smith, C.S., Children's Control of Some Complex Noun Phrases: A Repetition Study. University of Texas, unpublished paper, 1970.

Wechsler Preschool and Primary Scale of Intelligence, New York, New York, The Psychological Corporation, 1967.